

ARTICLE

Open Access

Genome sequencing broadens the range of contributing variants with clinical implications in schizophrenia

Bahareh A. Mojarad¹, Yue Yin¹, Roozbeh Manshaei², Ian Backstrom¹, Gregory Costain^{1,3}, Tracy Heung^{4,5}, Daniele Merico⁶, Christian R. Marshall⁷, Anne S. Bassett^{1,4,5,8} and Ryan K. C. Yuen^{1,9}

Abstract

The range of genetic variation with potential clinical implications in schizophrenia, beyond rare copy number variants (CNVs), remains uncertain. We therefore analyzed genome sequencing data for 259 unrelated adults with schizophrenia from a well-characterized community-based cohort previously examined with chromosomal microarray for CNVs (none with 22q11.2 deletions). We analyzed these genomes for rare high-impact variants considered causal for neurodevelopmental disorders, including single-nucleotide variants (SNVs) and small insertions/deletions (indels), for potential clinical relevance based on findings for neurodevelopmental disorders. Also, we investigated a novel variant type, tandem repeat expansions (TREs), in 45 loci known to be associated with monogenic neurological diseases. We found several of these variants in this schizophrenia population suggesting that these variants have a wider clinical spectrum than previously thought. In addition to known pathogenic CNVs, we identified 11 (4.3%) individuals with clinically relevant SNVs/indels in genes converging on schizophrenia-relevant pathways. Clinical yield was significantly enriched in females and in those with broadly defined learning/intellectual disabilities. Genome analyses also identified variants with potential clinical implications, including TREs (one in *DMPK*; two in *ATXN8OS*) and ultra-rare loss-of-function SNVs in *ZMYM2* (a novel candidate gene for schizophrenia). Of the 233 individuals with no pathogenic CNVs, we identified rare high-impact variants (i.e., clinically relevant or with potential clinical implications) for 14 individuals (6.0%); some had multiple rare high-impact variants. Mean schizophrenia polygenic risk score was similar between individuals with and without clinically relevant rare genetic variation; common variants were not sufficient for clinical application. These findings broaden the individual and global picture of clinically relevant genetic risk in schizophrenia, and suggest the potential translational value of genome sequencing as a single genetic technology for schizophrenia.

Introduction

Schizophrenia is a serious and disabling neuropsychiatric disorder that affects about 1% of the general population. Despite inherent heterogeneity, a century of research has provided strong evidence of genetic

predisposition, and statistical modelling has consistently indicated high heritability^{1,2}. However, discerning specific genetic risk factors for individuals with schizophrenia awaited technological advances in molecular genetics. Studies using first genome-wide chromosomal microarray (CMA) and then whole-exome sequencing (WES) have provided initial clues to the underlying genetic architecture of schizophrenia. These include contributions of rare (population frequency $\leq 0.1\%$) copy number variants (CNVs), other rare damaging and deleterious variants, common (population frequency $>1\%$) single-nucleotide

Correspondence: Anne S. Bassett (anne.bassett@utoronto.ca) or Ryan K. C. Yuen (ryan.yuen@sickkids.ca)

¹Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada

²Ted Rogers Centre for Heart Research, Cardiac Genome Clinic, The Hospital for Sick Children, Toronto, ON, Canada

Full list of author information is available at the end of the article

© The Author(s) 2021



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

polymorphisms (SNPs), and evidence for long-suspected polygenicity^{3–6}. Although the rare high-impact variants identified are often shared with other neurodevelopmental disorders (NDDs)^{7,8}, in contrast to autism spectrum disorder (ASD), intellectual disability (ID) and epilepsy, relatively few genetic findings for schizophrenia have reached the clinic^{1,9,10}.

Whole-genome sequencing (WGS) captures most forms of genetic variation across the genome in a single assay, surpassing the capabilities of CMA and WES combined¹¹. Furthermore, recent technical advances in WGS techniques and analyses allow for the genotyping of more complex genetic variation, such as tandem repetitive DNA elements, throughout the genome, not readily detectable using other sequencing techniques¹². The pathogenicity of large expansions of tandem DNA (in particular trinucleotide) repeats, has been extensively studied in over 40 genetic disorders, most of which are neurological but sometimes include psychosis¹³. Clinical observations of increased severity and/or younger age at the onset across successive generations historically suggested anticipation in schizophrenia, supporting the possible involvement of repeat expansions^{14,15}. However, the technologies and methodologies available to detect such repetitive DNA elements before now were limited.

In the current study, we applied WGS to a well-characterized community-based cohort of unrelated adults with schizophrenia. Our aim was, for the first time using a clinical lens and WGS data, to simultaneously detect multiple classes of genome-wide rare, high-impact genetic variants (including CNVs, single-nucleotide variants (SNVs), small insertions and deletions (indels), structural variants (SVs), and tandem repeat expansions (TREs), and assess for schizophrenia-related polygenic risk, while investigating possible phenotype correlations. Here, we defined high-impact variants as those with clinical relevance to schizophrenia or with potential clinical implications. Using this approach, we underscore the importance of thorough genome analyses in the identification of variants with potential clinical implications in individuals with schizophrenia, with or without molecular findings from routine CMA. This study thus expands on previous WGS studies of schizophrenia (Supplementary Table S1) to serve as an initial step in demonstrating the potential value of WGS as a single clinically relevant genetic technology for schizophrenia.

Methods

Study design

The 259 participants comprise a subset of a larger well-characterized cohort of unrelated adults who: (i) met standard diagnostic (DSM-5) criteria for schizophrenia or schizoaffective disorder, (ii) were of European descent, and (iii) were previously examined for the presence of rare

CNVs (≥ 10 kb in size) using CMA^{16,17}. Participants were ascertained from Canadian community mental health clinics and included individuals with schizophrenia across the IQ spectrum; details of the ascertainment strategy are described elsewhere^{16,17}. A priori, individuals with 22q11.2 microdeletions were excluded, as this established genetic subtype of schizophrenia is studied with WGS through a separate research initiative¹⁸. Also by design¹⁹, 136 (52.5%) of the individuals included in this study had broadly defined schizophrenia-relevant rare CNVs¹⁷ (Supplementary Tables S2, S3), 26 (10%) of whom had 28 CNVs previously classified as clinically relevant (pathogenic/likely pathogenic, Supplementary Table S2)^{16,17}. By including individuals with rare CNVs (10% with pathogenic CNVs), we undertook a conservative approach, interrogating for other potentially clinically relevant variants beyond well-studied CNVs.

Ethics statement

This study was approved by the Research Ethics Board at the Centre for Addiction and Mental Health (CAMH) (151/2002-02) and other local REBs^{16,17}. Written informed consent was obtained for all participants^{16,17}.

Assessment of the pathogenicity of rare variants (SNVs, indels, SVs, CNVs)

All rare (defined as population-based maximum allele frequency ≤ 0.01) exonic and exonic-splicing SNVs and indels, SVs, and CNVs were analyzed for their potential pathogenicity. Population allele frequency of each variant was derived from genomes included in ExAC, 1000 Genomes Project, gnomAD and gnomAD SV databases^{20–23}. Probability of loss-of-function (LoF) intolerance was measured by the upper bound of a Poisson-derived confidence interval around the ratio of the observed/expected number of LoF variants in every gene, derived from gnomAD (v2.1.1) and represented by LoF observed/expected upper bound fraction (LOEUF) score²⁰. LoF variants were defined as stop-gains, frameshift indels, and splice-site variants. Rare nonsynonymous variants with high predicted scores in 5 of 8 commonly used in silico algorithms [CADD (≥ 15), SIFT (≤ 0.05), PolyPhen2 HVAR (≥ 0.90), Provean (< -2.5), ma (≥ 1.90) and mt (≥ 0.5) scores, PhyloP Mam (≥ 2.30) and PhyloP Vert (≥ 4.0)] were considered as deleterious and were further assessed for pathogenicity²⁴. Given the evidence for genetic overlap between schizophrenia and other major NDDs, we conservatively considered only loci and genes as potentially associated with schizophrenia if they had been implicated in any NDD (e.g., ID or ASD), and their implicated pathways (Supplementary information)^{7,8,25–38}. Pathogenicity of rare SVs was assessed using their predicted damaging or deleterious effects on genes implicated in NDDs. In this study, for CNVs, SNVs and indels, we considered only

pathogenic and likely pathogenic variants as potentially clinically relevant²⁴ and contributing to the expression of schizophrenia, as adjudicated for NDDs.

Detection and independent confirmation of disease-associated tandem repeats using genome sequence data

To assess the presence of high-impact TREs in the genomes of our schizophrenia cohort, we collected data for 45 tandem repeat loci with known clinical associations, predominantly with neurological disorders (Supplementary Table S4). We used ExpansionHunter v3.0.2 to genotype these genomic repeat loci³⁹, and selected TREs larger than the described pathogenicity threshold for each locus for further characterization (Supplementary information). (<https://github.com/Illumina/ExpansionHunter>)⁴⁰. Rare TREs were classified as variants with a high-impact if the predicted size for the larger allele in each individual exceeded the disease-causing threshold for their loci¹³.

Clinical/demographic variables

We considered the following clinical/demographic variables for analyses: sex, presence or absence of family history of schizophrenia/psychotic illness, ID (broadly defined as borderline to moderate ID and non-verbal learning disability)¹⁶, syndromic features, and age at onset of schizophrenia (categorized as <18 years (“early”) or ≥18 years); details in Supplementary information, Supplementary Table S5, and as previously described^{16,17}. To assess these variables with respect to clinically relevant rare variant burden, we used a stringent definition, including only pathogenic CNVs and SNVs/indels, defined as above, but not TREs, following well-established guidelines²⁴.

Additional exploratory analyses

In addition to the primary focus on clinically relevant rare variants, we explored the possible role in our cohort of research-based genetic findings for schizophrenia, e.g., from exome sequencing and SNP-based studies.

Assessment of variants in putative schizophrenia-risk genes

To assess research-based SNV findings, we examined our cohort for all types of rare SNVs in ten genes reported to meet genome-wide significance for schizophrenia from recent meta-analysis results of exome sequencing data from the Schizophrenia Exome Sequencing Meta-analysis (SCHEMA) consortium (<https://schema.broadinstitute.org/>)⁴⁰.

Schizophrenia polygenic risk quantification

To assess the role of aggregate common variant background (polygenic risk score, PRS) for schizophrenia, we

used the training dataset provided by the 2014 PGC schizophrenia meta-analysis (Schizophrenia Working Group of the Psychiatric Genomics Consortium) to generate individual risk profile scores for our cohort³. There was no overlap of individuals in the training dataset with our schizophrenia cohort (Supplementary Fig. S1).

Non-psychiatric controls

We used comparable WGS data available from a previously published study of tetralogy of Fallot (TOF) and related congenital cardiac disease¹⁹ as a non-psychiatric control group to evaluate individual gene rare SNV findings and PRS results. After excluding seven genomes from individuals with TOF and a history of major neuropsychiatric conditions (e.g., ASD, psychotic mood disorder), data were available from 225 of the 232 individuals in this TOF cohort¹⁹ (Supplementary information).

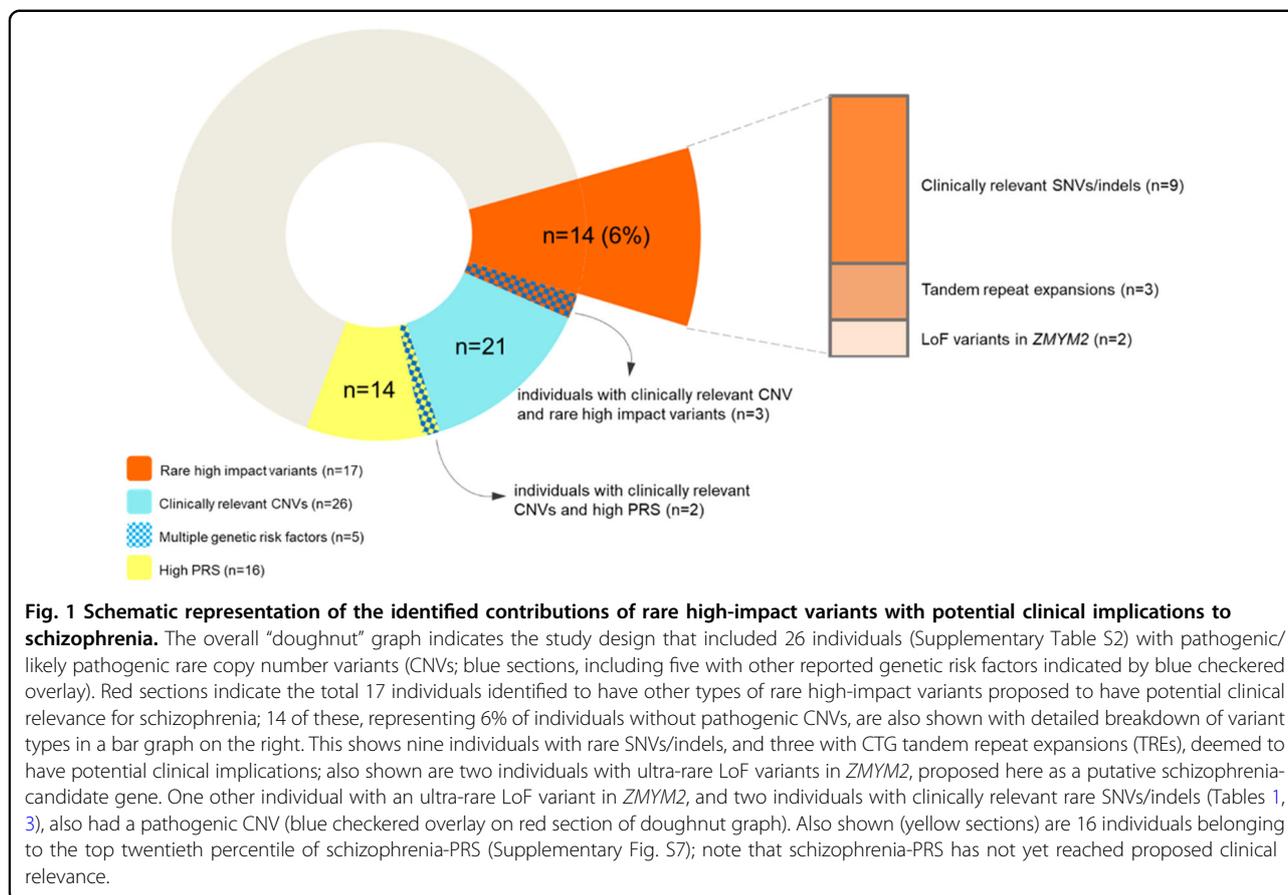
Results

Demographic and clinical features of the community-based cohort of 259 unrelated individuals with schizophrenia studied with genome sequencing are presented in Supplementary Table S5. The genomes sequenced had an average of 98.1% of bases covered by at least >1×, and an average mean depth of coverage of 38.42× (Supplementary Tables S6, S7). Restricting to exonic rare (allele frequency ≤1%) variants, we detected on average 271.7 SNVs, 20.1 indels, 2.89 SVs, and 3.33 CNVs (≥10Kb) per genome, consistent with expectations from previous WGS analyses of other samples⁴¹.

WGS enables simultaneous identification of multiple rare exonic variants of potential clinical relevance to schizophrenia

WGS identified several types of rare exonic variants of potential clinical relevance in this schizophrenia cohort. Importantly, the WGS pipeline identified 100% of the 28 rare CNVs in 26 individuals that were previously reported as clinically relevant (pathogenic/likely pathogenic)^{16,17} (Fig. 1, Supplementary Tables S2, S3). No rare SVs <10Kb met criteria as pathogenic/likely pathogenic.

In eleven individuals (4.3%), WGS also identified clinically relevant SNVs and indels with predicted deleterious effects on loss-of-function (LoF)-intolerant genes previously associated with schizophrenia-related NDDs (Methods and Table 1). There were five frameshift indels, three nonsense (stopgain), and three deleterious missense variants identified in ten genes: eight autosomal and two X-chromosome (both in females) (Table 1). Notably, two of the 11 individuals, both with missense variants, also had a clinically relevant 16p11.2 microduplication associated with increased schizophrenia risk (Table 1, Supplementary Tables S2 and S3)¹⁷. We thus propose clinically relevant SNVs/indels for nine (3.9%) of 233



individuals with no previously identified pathogenic CNVs (Fig. 1).

Rare disease-associated tandem repeat loci are expanded in schizophrenia

In three individuals, we identified and validated CTG TREs involving two of the 45 disease-associated tandem DNA repeats assessed. Two individuals had potentially damaging TREs in *ATXN8OS* and one had a potentially pathogenic TRE in *DMPK* (Table 2, Supplementary Fig. S2). The expanded CTG repeat (>200 repeats) at the 3' untranslated region (UTR) of *DMPK* was paternally inherited, with evidence of typical variable expression of the associated condition, myotonic dystrophy type 1 (DM1)⁴² (Table 2). Both expanded (>200 repeats) CTG repeats at the 3' UTR of *ATXN8OS* were found to be maternally inherited/derived (Table 2); there was no clinical or family history of typical neuromuscular features of the associated spinocerebellar ataxia type 8 (SCA8) disorder⁴³.

Therefore, of the 233 individuals with no pathogenic CNVs, we propose 14 individuals (6%) with rare high-impact variants (i.e., clinically relevant SNVs/indels, or TREs with clinical implications) (Fig. 1).

Females and individuals with learning/intellectual disabilities may have enhanced clinical yield from genome sequencing in schizophrenia

Individuals with learning and intellectual disabilities, as expected from previous studies of this cohort and other studies^{16,17,44}, were significantly enriched for rare clinically relevant CNV and/or SNV/indel variants ($p = 9.63 \times 10^{-6}$, Fig. 2). Results for clinically relevant variants were also significant for individuals with syndromic features ($p = 5.04 \times 10^{-5}$), and female sex ($p = 0.021$), but not for family history or age at onset (Fig. 2). Notably however, six (3.8%) of 158 individuals with no learning or intellectual disabilities (Supplementary Table S5) had a clinically relevant SNV/indel or disease-associated CTG TRE.

Additional findings relevant to the broader genetic architecture of schizophrenia

Rare clinically relevant SNVs/indels disrupt genes associated with neurodevelopmental pathways

The SNVs and indels identified to be of potential clinical relevance involve 10 genes associated with synaptic transmission ($n = 7$) or chromatin remodelling and transcription regulation ($n = 3$) (Table 1), consistent with

Table 1 Clinically relevant SNVs and indels in NDD-genes identified in eleven of 259 adults with schizophrenia.

Case	Sex	AAO (y)	FHx	ID/ syndromic	Rare CNV group	Clinically relevant CNV	Gene (transcript)	OMIM ID	Variant type	Variant	Chromosomal position (GRCh37/hg19)	Allele frequency (ExAC/ gnomAD)	Validation	Gene function
591	F	20	No	No/No	No	No	KCNQ5 (NM_001160130.1)	607357	Stopgain	c.C1984T: p.(Q662*)	chr6:73904349C>T	0/0	NA	Synaptic transmission
367	F	52	No	No/No	No	No	SCN8A (NM_001177984.2)	600702	Deletion (frameshift)	c.1940_1957delins: p.(G647Vfs*18)	chr12:52115634 GCGTGGTGTCCCTCATCG>GT	0/0	NA	
592	M	17	No	Mild/Yes	No	No	CACNA1A (NM_001127221.1)	601011	Deletion (frameshift)	c.2042_2043del: p.(Q681Rfs*100)	chr19:13414644 CCT>C	0/4.06 × 10 ⁻⁶	NA	
567	F	NA	Yes	Mild/No	No	No	SHANK3 (NM_033517.1)	606230	Deletion (frameshift)	c.2183_2184del: p.(Q728Qfs*12)	chr22:51153434 CAG>C	0/0	NA	
28	M	20	Yes	No/No	Yes ^{VUS}	No	RYR2 (NM_001035.2)	180902	Missense	c.C13489T: p.(R4497C)	chr1:237954741C>T	0/0	NA	
56	F	16	No	Mild/Yes	Yes	16p11.2 microduplication	SCN1B (NM_001037.4)	600235	Missense	c.C363G: p.(C121W)	chr19:935524558C>G	8.24 × 10 ⁻⁶ /4.06 × 10 ⁻⁶	Sanger ^{mat}	
55	F	31	Yes	Mild/Yes	Yes	16p11.2 microduplication	SYNJ1 (NM_133499.2)	313440	Missense	c.G1648A: p.(A550T)	chrX:47433735C>T	0/0	Sanger	
573	M	44	No	Mild/No	Yes ^{VUS}	No	MEIS2 (NM_001220482.1)	601740	Stopgain	c.C1099T: p.(Q367*)	chr15:37187379G>A	0/0	NA	Regulation of gene expression
625	F	24	No	Borderline/Yes	No	No	BRPFT (NM_001003694.1)	602410	Stopgain	c.C3346T: p.(R11116*)	chr3:9788005C>T	0/0	Sanger	
609	M	24	No	Borderline/Yes	No	No	BRPFT (NM_001003694.1)	602410	Insertion (frameshift)	c.2228dupA: p.(E743fs*5)	chr3:9784853G>GA	0/0	WES, DNV	
92	F	29	Yes	No/No	Yes ^{VUS}	No	MECP2 (NM_004992.3)	300005	Deletion (frameshift)	c.1157_1163del: p.(L386Hfs*21)	chrX:153296115 TGGGGCCA>T	0/0	Sanger	

Rare SNVs and indels considered clinically relevant, together with their key genetic parameters are shown. Two of these individuals also have clinically relevant CNVs (16p11.2 microduplications) F female, M male, AAO age at onset, y years, NA not available, FHx family history of schizophrenia/psychosis, ID borderline to mild intellectual disability, CNV copy number variant, VUS CNV of uncertain clinical significance, OMIM online mendelian inheritance in man, mat maternally inherited, WES whole-exome sequencing, DNV de novo variant. See Supplementary Table S3 for VUS CNVs.

Table 2 Three individuals with schizophrenia and TREs identified in the 3'UTR of genes *DMPK* and *ATXN80S*.

Gene	OMIM ID Disease	Case ID /family members	Sex	AAO (y)	FHx of SCZ	FHx of OMIM disease	ID	Other features	Rare CNV	CTG repeat	Chromosomal position (GRCh37/hg19)		Size	Validated	
											Normal	Pathogenic		EH-estimated	Validated
<i>DMPK</i>	160900 DM1	Proband 72	M	19	No	No	No	Episodes of severe constipation	Yes ^{VUS}	chr19:46273174-46273699	5-35	≥51	97	>200	
		Father	M	NA	NA	Yes	No	Cataract	NA				NA	>200	
<i>ATXN80S</i>	603680 SCA8	Mother	F	NA	NA	No	No	No	NA				NA	~9	
		Proband 329	M	21	Yes ^a	No	No	No	No	No	chr13:70713162-70713778	15-44	>110	116	>200
		Father	M	NA	NA	NA	No	No	No	NA				NA	~20
		Mother	F	NA	NA	No	No	Unilateral "glaucoma"	NA					NA	>200
<i>DMPK</i>	160900 DM1	Proband 11	F	17	No	Yes	Yes	Cataracts	Yes ^{VUS}				91	>200	
		Brother	M	NA	NA	NA	No	No	NA				NA	~20	
		Mother	F	NA	NA	No	Yes	No	Yes ^{VUS}				NA	~100	

TREs identified in *DMPK* and *ATXN80S* in three individuals with schizophrenia and their family members, together with their key clinical and genetic variables are shown
 F female, M male, AAO age at onset, y years, NA not applicable, FHx family history, SCZ schizophrenia/psychosis, ID borderline to mild intellectual disability, CNV copy number variant, OMIM online Mendelian inheritance in man EH ExpansionHunter; See Supplementary Table S4 for a list of tandem repeat loci examined in this study
^aProband 329 has a brother with schizophrenia; VUS, CNV of uncertain clinical significance, e.g., both proband 11 and her mother are identified to have a ~790 Kb deletion of uncertain significance overlapping *ZNF804A*; see Supplementary Table S3 for all VUS CNVs

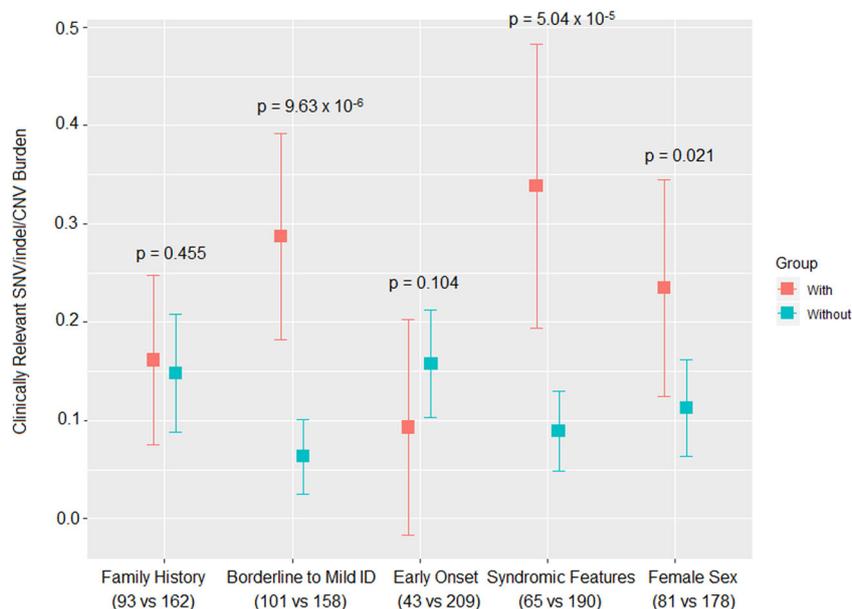


Fig. 2 Genetic risk for schizophrenia and clinical/demographic variables. This figure shows results for analyses of five clinical variables/features (family history of schizophrenia/psychosis, ID, early age at onset, mild syndromic features, and biological sex) with respect to rare clinically relevant variant burden, defined as the number of CNVs and/or SNVs/indels per individual. Orange and blue coloured boxes, and vertical bars representing 95% confidence intervals, indicate respectively results for individuals with and without each of the five variables; numbers for each subgroup are indicated in brackets under variable labels (Supplemental Table S5), and p -values for analyses are provided above graphed results. Clinically relevant rare variant burden was significantly greater for females, individuals with broadly defined ID, or with mild syndromic features.

pathways previously implicated in NDDs, including schizophrenia (Methods). Of the seven synaptic transmission genes, four harboured variants in genes encoding components of voltage-gated ion channels: *KCNQ5* p.(Q662*), *CACNA1A* p.(Q681Rfs*100), *SNC8A* p.(G647Vfs*18), and *SCN1B* p.(C121W), involving 1.5% ($n=4$) of the 259 individuals studied (Table 1). Of the genes involved in the regulation of gene expression, in two unrelated individuals, both with borderline ID and mild syndromic features³¹, we identified distinct ultra-rare (i.e., not seen in the general population) LoF variants affecting exons 13 and 7 of *BRPF1*, respectively a nonsense p.(R1116*) variant validated by Sanger sequencing and a confirmed de novo frameshift p.(E743fs*5) variant (Table 1, Supplementary Fig. S3).

Evidence for *ZMYM2* as a novel schizophrenia-candidate gene

In addition to *BRPF1*, we identified three other LoF-intolerant genes with multiple deleterious ultra-rare SNVs/indels in the schizophrenia cohort studied, and compared results to findings from previous studies of schizophrenia and other disorders (Table 3). The top candidate gene identified was *ZMYM2* with three rare LoF variants in three unrelated individuals ($p=9.51 \times 10^{-6}$); one had a *NRXN1* deletion (Fig. 1, Table 3, Supplementary Table S2). *ZMYM2* was supported by

substantial evidence from the literature^{45–47} including suggestive meta-analysis results from exome studies of schizophrenia (SCHEMA p -value = 1.79×10^{-5})⁴⁰, and no rare LoF variants detected in our TOF-control sample (Table 3, Supplementary Fig. S4). Supplementary Table S8 shows the LoF and rare missense variants in *ZMYM2* identified in our cohort and reported by others. Though interesting genes with evidence of greater constraint, i.e., lower gnomAD LoF observed/expected upper bound fraction (LOEUF), variant results for *GPRIN1* and *DNAJC6* were less compelling as candidate genes of potential clinical relevance (Table 3).

Contribution of other rare exonic variants and polygenic risk

Using our WGS data we examined the ten genes showing genome-wide significant association with schizophrenia on meta-analyses using SCHEMA exome sequencing data⁴⁰. In these ten genes we detected between one and 12 (in *SETD1A*) rare missense exonic variants, none of which were considered clinically relevant (Supplementary Table S9).

Analyses using common variants showed that our schizophrenia cohort had a significantly higher mean PRS compared to the TOF-control group, explaining 9.5% of the variance (Nagelkerke's pseudo R^2 from logistic regression at $P_T=0.05$, $p=1.07 \times 10^{-9}$) (Supplementary Fig. S5). Mean PRS was not significantly different between those

Table 3 Genes with deleterious ultra-rare LoF variants identified in two or more individuals with schizophrenia in the sample studied.

Gene (gnomAD LOEUF)	This schizophrenia sample (total n = 259)		Results from other studies				Gene function			
	Number of individuals with variants in gene	Variant	Case	Other clinically relevant variant	Binomial p-value	FDR		Number of LoF variants ^a	References	SCHEMA meta-analysis p-value ^b
ZMYM2 (0.33)	3	c.G1960T: p.(E654*) c.2253dupT: p.(S751Rfs*2) c.763_766del: p.(T343Lfs*37)	562 ^{rare} 183 ^{rare} 291	2p16.3 <i>NRXN1</i> deletion No No	9.51x10 ⁻⁶	0.028	7 SCZ, 2 ASD, 7 ID/DD	45-47/57	1.79 x 10 ⁻⁵	Regulation of gene expression
GPRIN1 (0.32)	2	c.2600_2604del: p.(E867Afs*40) c.1334_1335del: p.(V445Gfs*27)	34 ^{rare} 16 ^{rare}	No No	9.08x10 ⁻⁵	0.134	1 SCZ	45	1	Synaptic development
BRPF1 (0.18) (also in Table 1)	2	c.C3346T: p.(R1116*) c.2228dupA: p.(E743fs*5)	625 609 ^c	No No	3.67x10 ⁻⁴	0.360	3 SCZ	645	0.017	Regulation of gene expression
DNAJC6 (0.22)	2	c.C988T: p.(R330*) c.94_95del: p.(S32Rfs*49)	585 26	No No	7.39x10 ⁻⁴	0.544	0	-	0.527	Vesicle mediated transport

Binomial p-value and FDR were calculated as described in Supplementary Information. Ultra-rare: not seen in the general population

LoF loss of function, FDR false discovery rate, gnomAD genome aggregation database, LOEUF loss-of-function observed/expected upper bound fraction, SCZ schizophrenia, ASD autism spectrum disorder, ID/DD intellectual disability/developmental delay

^aThe number of previously reported LoF variants in NDDs is indicated for each gene and condition associated. See Supplementary Table S8 and Supplementary Fig. S4 for details of ZMYM2 variants in this and other cohorts.

^bSuperscript ^{rare} indicates an individual belonging to the rare CNV subgroup (see Supplementary Table S3)

^cGenome-wide p-value reported in SCHEMA exome-sequencing meta-analysis (schema.broadinstitute.org/); only the p value for ZMYM2 met criteria for genome-wide suggestive

^dIn this individual, the LoF variant in BRPF1 was confirmed as a de novo variant (see Table 1, Supplementary Fig. S3)

with and without clinically relevant rare variants ($p = 0.52$) (Supplementary Fig. S6a). Sixteen (6.23%) individuals with schizophrenia (Fig. 1), fell in the top twentieth PRS percentile subgroup (i.e., where the odds ratio (OR) was greatest relative to the remainder of the sample, $OR = 2.92$, 95% confidence interval: 1.05–8.11) (Supplementary Fig. S7). Individuals with a positive family history of schizophrenia/psychoses showed a higher mean PRS than those without such a family history ($p = 0.046$); no other results for clinical/demographic variables achieved significance though there was a non-significant trend for higher mean PRS in individuals with no learning and/or intellectual disabilities ($p = 0.085$) (Supplementary Fig. 6b).

Per our original study design^{16,17}, we also compared findings between the 136 individuals with rare CNVs, including those of uncertain clinical significance (all identified by the WGS pipeline; Supplementary Tables S2, S3), to the 123 individuals with no rare CNVs. There were no significant differences, respectively, for the 11 rare SNVs/indels ($n = 5$ vs $n = 6$, Fisher's exact test, $p = 0.76$), global burden of ultra-rare LoF variants (one-sided Wilcoxon signed-rank test, $p = 0.0975$, data not shown), or mean PRS ($p = 0.37$) (Supplementary Fig. 6a).

Discussion

Using genome sequences from 259 unrelated schizophrenia-affected adults, and simultaneously interrogating for a range of genetic variants, we undertook a conservative approach to explore the clinical relevance of other types of variants to schizophrenia beyond copy number variation. We considered only rare SNVs and indels with a strong association with NDDs and determined that about 4% of the studied individuals had such clinically relevant, predominantly LoF, variants. In addition, we identified trinucleotide TREs with potential clinical implications in other individuals. The fact that 14 (5.4%) of 259 individuals with schizophrenia, including many with no broadly defined learning disability/ID, had a high-impact SNV/TRE not detectable by CMA, provides an initial indication that an important minority of patients in the community would be found to have such clinically relevant variants using WGS. Notably, several individuals had multiple genetic risk factors (Fig. 1, Table 1), consistent with polygenicity within individuals, and reduced penetrance, even of high-impact clinically relevant variants in schizophrenia^{17,48,49}.

A novel finding was the identification of individuals with schizophrenia and TREs associated with DM1 and SCA8, neuromuscular disorders with highly variable expressivity and neuropsychiatric manifestations^{15,42,43}. A recent genome-wide study of >17,200 individuals identified rare CTG TREs in *DMPK* in individuals with ASD⁵⁰, and older studies reported high prevalence of psychotic disorders in individuals with TREs in *DMPK* and juvenile

DM1⁵¹. Older technologies had provided initial evidence linking CTG repeats in *ATXN80* and *SCA8* to major psychoses¹⁵, and overall TRE results may be to some extent consistent with other historical studies^{14,52–54}. With continuous technical improvements in genome-wide detection of such expanded repeats and more precise size estimations, we expect to identify other, novel unstable TREs in individuals with schizophrenia, and elucidate the underlying mechanisms that lead to psychiatric expression.

The fact that the clinically relevant rare SNVs/indels identified involved genes implicated in synaptic transmission and transcription regulation pathways previously associated with neurodevelopment and schizophrenia^{6,8,37}, further illustrates the potential for WGS to contribute to etiological understanding with convergence on mechanisms of importance to schizophrenia pathogenesis. In an iterative fashion, each genome can potentially inform future adjudication of other clinically relevant variants, and can be reanalyzed as further discoveries are made⁵⁵.

The clinical/demographic findings suggest that not only individuals with any degree of intellectual impairment but also female patients may disproportionately benefit from clinical genetic testing in schizophrenia¹⁶. The latter finding is consistent with the possibility that there may be in the general population a female protective mechanism for schizophrenia analogous to that proposed in other NDDs with male bias of expression, such as ASD⁵⁶. Studies of larger schizophrenia cohorts are needed to investigate this phenomenon.

In addition to findings of possible immediate impact to clinical translation, the sequencing data allowed us to examine several other pertinent aspects of genetic architecture in schizophrenia. This includes our proposal of Zinc Finger MYM-Type Containing 2 (*ZMYM2*) (OMIM: 602221) as a putative schizophrenia-risk candidate gene of possible clinical relevance, supported by a gene-based analysis of deleterious ultra-rare variants in LoF-intolerant genes, SCHEMA meta-analysis data⁴⁰, and LoF variants in *ZMYM2* reported in other studies examining NDDs, including ASD, ID and schizophrenia^{45–47,57}. Comparable to other schizophrenia-associated genes, variants in *ZMYM2* appear to manifest pleiotropic effects^{45–47,57}. *ZMYM2* encodes a zinc finger protein, which may act as a transcription factor and thereby regulate gene expression (Supplementary Fig. S4), a mechanism elucidated for other clinically relevant variants (Table 1). Further studies examining the function of the encoded protein during neurodevelopment, and analysis of larger psychiatric cohorts, are required to establish a robust link between LoF variants in *ZMYM2* and schizophrenia pathogenesis. Functional studies and further sequencing data will also be needed for more broadly defined rare nonsynonymous

variants, given their expected lower impact than ultra-rare LoF variants. This would include results for genes proposed through exome-sequencing and meta-analysis (e.g., SCHEMA consortium; Supplementary Table S9)⁴⁰, in order to determine their clinical relevance and contribution to overall schizophrenia liability.

We also took advantage of WGS data to simultaneously assess the potential contribution of common variant burden (e.g., explaining an estimated 9.5% of the variance using PRS data) and rare clinically relevant variants, allowing for examination of a broader-based risk profile for each individual, consistent with the proposed polygenic nature of schizophrenia. Unlike previous studies that used imputed genotyping data to study PRS⁵⁸, here we implemented precisely genotyped SNP data as determined by WGS (not possible using WES or CMA). We did not identify a correlation between PRS and the burden of rare, clinically relevant variants (Supplementary Fig. S6a)¹⁸. While this may in part have been affected by the exclusion from this cohort of individuals with one of the highest known risks for schizophrenia, 22q11.2 deletions, there are as yet limited data on PRS in the context of other high-impact variants associated with schizophrenia¹⁸. Consistent with other studies⁵⁸, individuals with a family history of schizophrenia/psychosis showed some enrichment for schizophrenia PRS, (supplementary information and Supplementary Fig. S6b), but limited availability of parental DNA samples precluded variant segregation analyses to confirm the transmission of SNP-based risk alleles. Judging by the modest estimated risk conveyed by the highest PRS (OR = 2.92; 95% CI: 1.05-8.11), and consistent with other reports⁵⁹, PRS is not yet sufficient to apply clinically for individual schizophrenia risk classification. Nevertheless, the growing numbers of individuals with available WGS data deserve further study of the potential clinical application of polygenic risk prediction, and effects on this of high-impact rare variants^{18,60}.

Our results should be interpreted in the context of a few important limitations. First, by design, and comparable to a companion study of TOF¹⁹, about half of this cohort had rare genic CNVs, as determined by previous CMA analysis^{16,17}. The majority of such CNVs are not clinically pathogenic (Supplementary Tables S2, S3), nevertheless, this may have influenced our findings, including the clinical yield of SNVs, TREs and SVs, and results for clinical variables. Analysis of larger cohorts of schizophrenia with and without clinically relevant CNVs would be needed to more precisely estimate their impact relative to other high-impact variants and polygenic risk, and relationships to clinical phenotypes^{18,61}. Second, due to technical limitations and the complex nature of tandem repeats, we could not determine the precise size of TREs³⁹. Size underestimation may have hindered the detection of other tandem repeats contributing to

schizophrenia risk in this cohort. Third, inability to identify schizophrenia-relevant LoF SVs may have been in part due to the limited resources currently available for the interpretation of such variants. Efforts are underway to construct comprehensive resources for SVs, which together with the improvements in our understanding of the complex etiology and genetic mechanisms of schizophrenia will enhance the identification of phenotypically relevant variants through WGS-based approaches. Fourth, using a cohort of 225 adults with congenital cardiac disease as controls might have produced more conservative results for our analyses than if using other control groups without a developmental, albeit cardiac, phenotype. However, there are few known links between genetic risk for TOF and for schizophrenia (apart from e.g., 22q11.2 deletions, 1q21.1 duplications, and deleterious variants in *RYR2* (Supplementary information)). Fifth, our clinical adjudication of variants could also be considered overly conservative, relying on currently available results for NDD, and 45 established genes for TREs. Future efforts to refine clinical interpretation of rare variants for schizophrenia will be essential.

In conclusion, our results provide important evidence of the enhanced performance of WGS compared to CMA in the detection of genome-wide clinically relevant variants⁶², and an initial indication of features that could help identify individuals with schizophrenia who are most likely to benefit from clinical genetic testing and genetic counseling^{16,17}. The results also reiterate the complexity and pleiotropy of schizophrenia, and suggest the interplay of multiple variant types, each with varied expressivity and penetrance, in every individual. With continued improvements in high-throughput sequencing technologies, WGS will become more affordable, which together with advances in interpretation (particularly for variants affecting non-coding and regulatory elements) promise to make WGS an ideal tool for routine diagnostic practice⁶³. Global efforts combining WGS data from various neuropsychiatric disorders will shed light on the shared and disparate genetic factors and mechanisms underlying these disorders⁵⁰. Eventually, implementation of clinical WGS will extend to patients with schizophrenia, as for those with other NDDs, to further guide our understanding of prognosis, medical management, and familial recurrence risk assessment, and as part of global efforts towards “precision medicine”.

Acknowledgements

The authors are grateful to all patients and their families for their participation in this study. We thank The Centre for Applied Genomics (TCAG, a node of CGEn), which is supported by the Canada Foundation of Innovation, Genome Canada, the Hospital for Sick Children, and partners. We thank Dr. Nozomu Sato for designing primers for TRE validation in *DMPK*. We also thank Wilson Sung, Bhooma Thiruvahindrapuram, Dr. John Wei, Dr. Miriam S. Reuter, Dr. Robert Davies and Dr. Stephen W Scherer at TCAG for their helpful discussions on PRS analysis and WGS quality assessment, and Dr. Chelsea Lowther and many research assistants, trainees, and clinical colleagues for their efforts that

were essential to study recruitment and data collection. This work was supported by the Canadian Institutes of Health Research (CIHR) (MOP-89066 to A.S.B., MOP-111238 to A.S.B.), and a Canada Research Chair in Schizophrenia Genetics and Genomic Disorders (Tier 1, 2009-2016 to A.S.B.). A.S.B. holds the Dalglish Chair in 22q11.2 Deletion Syndrome at the University Health Network and University of Toronto. This work was also funded by a Nancy E.T. Fahrner Award and Catalyst Scholar in Genetics from The Hospital for Sick Children to R.K.C.Y., and with support from the University of Toronto McLaughlin Centre and the Hospital for Sick Children Foundation.

Author details

¹Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada. ²Ted Rogers Centre for Heart Research, Cardiac Genome Clinic, The Hospital for Sick Children, Toronto, ON, Canada. ³Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, ON, Canada. ⁴Clinical Genetics Research Program, Centre for Addiction and Mental Health, Toronto, ON, Canada. ⁵The Dalglish Family 22q Clinic for Adults with 22q11.2 Deletion Syndrome, Toronto General Hospital, University Health Network, Toronto, ON, Canada. ⁶Deep Genomics Inc., Toronto, Ontario and The Centre for Applied Genomics (TCAG), The Hospital for Sick Children, Toronto, ON, Canada. ⁷Paediatric Laboratory Medicine, Genome Diagnostics, The Hospital for Sick Children, Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada. ⁸Department of Psychiatry, University of Toronto, Toronto General Hospital Research Institute and Campbell Family Mental Health Research Institute, Toronto, ON, Canada. ⁹Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada

Author contributions

B.A.M., R.K.C.Y. and A.S.B. conceived and coordinated the project. B.A.M. processed and analyzed the whole genome sequencing data. Y.Y. processed PRS and ethnicity, and with T.H. the genotype-phenotype correlation and analyses. B.A.M. and I.B. designed and performed experiments for variant validation. R.M. analyzed ultra-rare variants with contributions from D.M. B.A.M. interpreted the clinically relevant variants with contributions from C.R.M. and G.C. A.S.B. managed, recruited, diagnosed and with G.C. examined the recruited participants. B.A.M., R.K.C.Y. and A.S.B. wrote the manuscript with contributions from all authors.

Conflict of interest

Daniele Merico is a full-time employee of Deep Genomics Inc. and is entitled to a stock option. All other authors report no financial relationships with commercial interests.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41398-021-01211-2>.

Received: 28 June 2020 Revised: 28 December 2020 Accepted: 12 January 2021

Published online: 01 February 2021

References

- Costain, G. & Bassett, A. S. Clinical applications of schizophrenia genetics: genetic diagnosis, risk, and counseling in the molecular era. *Appl. Clin. Genet.* **5**, 1–18 (2012).
- Cannon, T. D., Kaprio, J., Lönqvist, J., Huttunen, M. & Koskenvuo, M. The genetic epidemiology of schizophrenia in a Finnish twin cohort: a population-based modeling study. *JAMA Psychiatry* **55**, 67–74 (1998).
- Schizophrenia Working Group of the Psychiatric Genomics Consortium. et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421 (2014).
- Marshall, C. R. et al. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat. Genet.* **49**, 27–35 (2017).
- International Schizophrenia Consortium. et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
- Xu, B. et al. De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nat. Genet.* **44**, 1365 (2012).
- Rees, E. et al. Analysis of intellectual disability copy number variants for association with schizophrenia. *JAMA Psychiatry* **73**, 963–969 (2016).
- McCarthy, S. E. et al. De novo mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. *Mol. Psychiatry* **19**, 652 (2014).
- Miller, D. T. et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am. J. Hum. Genet.* **86**, 749–764 (2010).
- Poduri, A., Sheidley, B. R., Shostak, S. & Ottman, R. Genetic testing in the epilepsies-developments and dilemmas. *Nat. Rev. Neurol.* **10**, 293–299 (2014).
- Clark, M. M. et al. Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. *npj Genom. Med.* **3**, 16 (2018).
- Alkan, C., Coe, B. P. & Eichler, E. E. Genome structural variation discovery and genotyping. *Nat. Rev. Genet.* **12**, 363–376 (2011).
- Pearson, C. E., Edamura, K. N. & Cleary, J. D. Repeat instability: mechanisms of dynamic mutations. *Nat. Rev. Genet.* **6**, 729–742 (2005).
- Bassett, A. S. & Honer, W. G. Evidence for anticipation in schizophrenia. *Am. J. Hum. Genet.* **54**, 864–870 (1994).
- Vincent, J. B. Unstable repeat expansion in major psychiatric disorders: two decades on, is dynamic DNA back on the menu? *Psychiatr. Genet.* **26**, 156–165 (2016).
- Lowther, C. et al. Impact of IQ on the diagnostic yield of chromosomal microarray in a community sample of adults with schizophrenia. *Genome Med.* **9**, 105–105 (2017).
- Costain, G. et al. Pathogenic rare copy number variants in community-based schizophrenia suggest a potential role for clinical microarrays. *Hum. Mol. Genet.* **22**, 4485–4501 (2013).
- Clynen, I. et al. Genetic contributors to risk of schizophrenia in the presence of a 22q11.2 deletion. *Mol. Psychiatry* <https://doi.org/10.1038/s41380-020-0654-3> (2020 e-published).
- Reuter, M. S. et al. Haploinsufficiency of vascular endothelial growth factor related signaling genes is associated with tetralogy of Fallot. *Genet. Med.* **21**, 1001–1007 (2019).
- The Genome Aggregation Database (gnomAD). <https://gnomad.broadinstitute.org/>.
- The Exome Aggregation Consortium (ExAC). <https://exac.broadinstitute.org/>.
- The International Genome Sample Resource (1000 Genomes Project). <https://www.internationalgenome.org/>.
- Collins, R.L. et al. An open resource of structural variation for medical and population genetics. *bioRxiv* <https://doi.org/10.1101/578674> (2019).
- Richards, S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405 (2015).
- Guilmatre, A. et al. Recurrent rearrangements in synaptic and neurodevelopmental genes and shared biologic pathways in schizophrenia, autism, and mental retardation genes. *JAMA Psychiatry* **66**, 947–956 (2009).
- Sebat, J., Levy, D. L. & McCarthy, S. E. Rare structural variants in schizophrenia: one disorder, multiple mutations; one mutation, multiple disorders. *Trends Genet.* **TIG 25**, 528–535 (2009).
- Damaj, L. et al. CACNA1A haploinsufficiency causes cognitive impairment, autism and epileptic encephalopathy with mild cerebellar symptoms. *Eur. J. Hum. Genet.* **23**, 1505–1512 (2015).
- O'Brien, J. E. & Meisler, M. H. Sodium channel SCN8A (Nav1.6): properties and de novo mutations in epileptic encephalopathy and intellectual disability. *Front. Genet.* **4**, 213 (2013).
- Carney, R. M. et al. Identification of MeCP2 mutations in a series of females with autistic disorder. *Pediatr. Neurol.* **28**, 205–211 (2003).
- Lehman, A. et al. Loss-of-function and gain-of-function mutations in KCNQ5 cause intellectual disability or epileptic encephalopathy. *Am. J. Hum. Genet.* **101**, 65–74 (2017).
- Yan, K. et al. Mutations in the chromatin regulator gene BRPF1 cause syndromic intellectual disability and deficient histone acetylation. *Am. J. Hum. Genet.* **100**, 91–104 (2017).

32. Fassio, A. et al. SYN1 loss-of-function mutations in autism and partial epilepsy cause impaired synaptic function. *Hum. Mol. Genet.* **20**, 2297–2307 (2011).
33. Wallace, R. H. et al. Febrile seizures and generalized epilepsy associated with a mutation in the Na⁺-channel β 1 subunit gene SCN1B. *Nat. Genet.* **19**, 366 (1998).
34. Durand, C. M. et al. Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nat. Genet.* **39**, 25 (2006).
35. Giliberti, A. et al. MEIS2 gene is responsible for intellectual disability, cardiac defects and a distinct facial phenotype. *Eur. J. Med. Genet.* (2019).
36. Ambalavanan, A. et al. De novo variants in sporadic cases of childhood onset schizophrenia. *Eur. J. Hum. Genet.* **24**, 944 (2015).
37. Fromer, M. et al. De novo mutations in schizophrenia implicate synaptic networks. *Nature* **506**, 179–184 (2014).
38. Tsankova, N., Renthal, W., Kumar, A. & Nestler, E. J. Epigenetic regulation in psychiatric disorders. *Nat. Rev. Neurosci.* **8**, 355 (2007).
39. Dolzhenko, E. et al. ExpansionHunter: a sequence-graph-based tool to analyze variation in short tandem repeat regions. *Bioinformatics* **35**, 4754–4756 (2019).
40. Tajinder Singh, T.P., Curtis, D., Akil, H., Neale, B. M. & Daly M. J. On behalf of the Schizophrenia Exome Meta-Analysis (SCHEMA) Consortium. Exome sequencing identifies rare coding variants in 10 genes which confer substantial risk for schizophrenia. *medRxiv* <https://doi.org/10.1101/2020.09.18.20192815> (2020).
41. Yuen, R. K. et al. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat. Neurosci.* **20**, 602 (2017).
42. Harper, P. S. *Myotonic Dystrophy* (Oxford University Press, 2009).
43. Koob, M. D. et al. An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). *Nat. Genet.* **21**, 379–384 (1999).
44. Khandaker, G. M., Barnett, J. H., White, I. R. & Jones, P. B. A quantitative meta-analysis of population-based studies of premorbid intelligence and schizophrenia. *Schizophrenia Res.* **132**, 220–227 (2011).
45. Genovese, G. et al. Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat. Neurosci.* **19**, 1433 (2016).
46. Howrigan, D. P. et al. Exome sequencing in schizophrenia-affected parent-offspring trios reveals risk conferred by protein-coding de novo mutations. *Nat. Neurosci.* **23**, 185–193 (2020).
47. Neale, B. M. et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* **485**, 242 (2012).
48. Lowther, C. et al. Molecular characterization of NRXN1 deletions from 19,263 clinical microarray cases identifies exons important for neurodevelopmental disease expression. *Genet. Med.* **19**, 53–61 (2017).
49. Vassos, E. et al. Penetrance for copy number variants associated with schizophrenia. *Hum. Mol. Genet.* **19**, 3477–3481 (2010).
50. Trost, B. et al. Genome-wide detection of tandem DNA repeats that are expanded in autism. *Nature* **586**, 80–86 (2020).
51. Douniol, M. et al. Psychiatric and cognitive phenotype of childhood myotonic dystrophy type 1. *Dev. Med. Child Neurol.* **54**, 905–11 (2012).
52. O'Donovan, M. C. et al. Confirmation of association between expanded CAG/CTG repeats and both schizophrenia and bipolar disorder. *Psychol. Med.* **26**, 1145–1153 (1996).
53. Lindblad, K. et al. Detection of expanded CAG repeats in bipolar affective disorder using the repeat expansion detection (RED) method. *Neurobiol. Dis.* **2**, 55–62 (1995).
54. McInnis, M. G. et al. Anticipation in bipolar affective disorder. *Am. J. Hum. Genet.* **53**, 385–390 (1993).
55. Costain, G. et al. Periodic reanalysis of whole-genome sequencing data enhances the diagnostic advantage over standard clinical genetic testing. *Eur. J. Hum. Genet.* **26**, 740–744 (2018).
56. Werling, D. M. & Geschwind, D. H. Sex differences in autism spectrum disorders. *Curr. Opin. Neurol.* **26**, 146–153 (2013).
57. Stessman, H. A. F. et al. Targeted sequencing identifies 91 neurodevelopmental disorder risk genes with autism and developmental disability biases. *Nat. Genet.* **49**, 515–526 (2017).
58. Agerbo, E. et al. Polygenic risk score, parental socioeconomic status, family history of psychiatric disorders, and the risk for schizophrenia: a Danish population-based study and meta-analysis. *JAMA Psychiatry* **72**, 635–641 (2015).
59. Bogdan, R., Baranger, D. A. A. & Agrawal, A. Polygenic risk scores in clinical psychology: bridging genomic risk to individual differences. *Annu. Rev. Clin. Psychol.* **14**, 119–157 (2018).
60. Khera, A. V. et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).
61. Tansey, K. E. et al. Common alleles contribute to schizophrenia in CNV carriers. *Mol. Psychiatry* **21**, 1085 (2015).
62. Stavropoulos, D. J. et al. Whole genome sequencing expands diagnostic utility and improves clinical management in pediatric medicine. *NPJ Genom. Med.* **1**, 15012 (2016).
63. Prokop, J. W. et al. Genome sequencing in the clinic: the past, present, and future of genomic medicine. *Physiol. Genomics* **50**, 563–579 (2018).